

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04N 7/01	A1	(11) International Publication Number: WO 96/41469 (43) International Publication Date: 19 December 1996 (19.12.96)
(21) International Application Number: PCT/US96/09813 (22) International Filing Date: 7 June 1996 (07.06.96) (30) Priority Data: 08/488,222 7 June 1995 (07.06.95) US (71)(72) Applicant and Inventor: GESHWIND, David, M. [US/US]; 111 Fourth Avenue, New York, NY 10003 (US). (74) Agent: STEIN, Mitchell, A.; Stein & Associates, P.C., 805 Third Avenue, New York, NY 10022 (US).		(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>

(54) Title: SYSTEMS USING MOTION DETECTION, INTERPOLATION, AND CROSS-DISSOLVING FOR IMPROVING PICTURE QUALITY

- | |
|---|
| (1001) IMAGE ANALYSIS ALGORITHMS EXTRACT 3D SHAPE AND MOTION |
| (1002) BITMAPS EXTRACTED FROM OBJECTS AND "UNWRAPPED" BY APPLYING INVERSE PROJECTIVE GEOMETRY TRANSFORMS |
| (1003) 3D SHAPE AND MOTION DATA USED TO PRODUCE INTERMEDIATE 3D "SCENES" WITH REPOSITIONED/RESHAPED OBJECTS |
| (1004) TEXTURE MAPS SELECTED OR CROSS-DISSOLVED BETWEEN |
| (1005) TEXTURE MAPS REAPPLIED TO NEWLY CREATED 3D OBJECTS |

Software Flow Diagram of 3D Interpolation Embodiment

(57) Abstract

Image sequences are analyzed and, with optional human assistance, computer calculated shape or motion data (2D or 3D) is extracted (1001). That data is applied to distort pairs of source frames (1002) and, along with cross-dissolving between distorted pairs, intermediate image frames are thereby produced (1003). The resulting higher-frame rate signal consists of original and intermediate frames with each composed and positioned appropriately in time, and with each comprising unique visual information. Improved visual quality or, alternately, data reduction is thereby achieved. In the latter case, the vacated bandwidth may be used for: additional resolution or definition (in both directions or in bit-depth); additional image area (e.g. HDTV, widescreen or "letterbox" side-strips); 3D information in the form of a second image, or from which two images can be created by combination with the first; interactive or game data; hyper- or multimedia data; image segmentation data showing areas of motion or where different algorithms are to be applied; or the interleaving of several program channels.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

SYSTEMS USING MOTION DETECTION, INTERPOLATION, AND CROSS-DISSOLVING FOR IMPROVING PICTURE QUALITY

5

TECHNICAL FIELD

The instant invention comprises a method, process or algorithm, and variations thereon, which method includes motion detection, cross-dissolving and shape interpolation; devices or systems for practicing that method; and, product (generally motion picture film, videotape or videodisc, analog or digitally stored motion sequences on magnetic or optical media, or a transmission, broadcast or other distribution of same) produced by the method and/or system.

SCOPE OF INVENTION AND PRIOR ART

The instant invention comprises a method, process or algorithm, and variations thereon, including motion detection, cross-dissolving and shape interpolation; devices or systems for practicing that method; and, product (generally motion picture film, videotape or videodisc, analog or digitally stored motion sequences on magnetic or optical media, or a transmission, broadcast or other distribution of same) produced by the method and/or system.

The purpose to which the invention is applied is to process (generally by digital computer image processing) a motion picture sequence in order to produce a processed motion picture sequence which exhibits: an increase in the perceived quality of that sequence when viewed; and/or a decrease of the requirements for information storage or transmission resources without significantly effecting image quality (i.e., data compression or bandwidth reduction).

In order to accomplish these benefits, Inventor will be relying on a number of methods and devices that are well-known, well-developed, well-documented and within the ken of intended practitioners and those skilled in the art.

The intended practitioner of the present invention is someone who is skilled in designing, implementing, integrating, building, creating, programming or utilizing processes, devices, systems and products, such as those that: encode a higher-definition television or video signal into a lower-definition television or video signal suitable for transmission, display or recording; record, transmit, decode or display such an encoded signal; transduce or transfer an image stream from an imaging element to a transmission or storage element, such as a television camera or film chain; transfer an image stream from a signal input to a recording medium, such as a videotape or videodisc recorder; transfer an image stream from a recording medium to a display element, such as a videotape or videodisc player; transfer data representing images from a computer memory element to a display element, such as a framestore or frame buffer; synthesize an image output stream from a mathematical model, such as a computer graphic rendering component; modify or combine image streams, such as image processing components, time-base correctors, signal processing components, or special effects components; products that result from the foregoing; and many other devices, processes and products that fall within the realms of motion picture and television engineering, or computer graphics and image processing.

That is, one skilled in the art required to practice the instant invention is capable of one or more of the following: design and/or construction of devices, systems, hardware and software (i.e., programming) for motion picture and television production, motion picture and television post production, signal processing, image processing, computer graphics, and the like. That is, motion picture and television engineers, computer graphic system designers and programmers, image processing system designers and programmers, digital software and hardware engineers,

communication and information processing engineers, applied mathematicians, etc.

Those skilled in the art know how to accomplish such tasks as to: design and construct devices, design and integrate systems, design software for and program those devices and systems, and utilize those devices and systems to create information product, which devices and systems transfer and/or transform information derived from image streams. Further, such practitioners are skilled in providing "software glue"; that is to take known or existing algorithms, programs, utilities, subroutines and libraries and to take the output from one such program and direct it to the input of another. Sometimes that task requires that the output data be manipulated or reformatted prior to its use as input, and such file and data conversion is also within the skill in the art. Such processes, programs, devices and systems comprise well known digital or analog electronic hardware, and software, components. The details of accomplishing such standard tasks are well known and within the ken of those skilled in these arts; are not (in and of themselves) within the scope of the instant invention; although some novel details of implementation, new uses and new systems designs are. These known elements will be referred to but not described in detail in the instant disclosure.¹

Rather, what will be disclosed are novel and high-level: image analysis and processing algorithms; information flows; and, system designs. Disclosed will be what one skilled in the art will need to know, beyond that with which he is already familiar, in order to implement the instant invention. These algorithms and system designs will be presented by description, algebraic formulae and graphically, as is standard and frequent practice in the fields of motion picture and television engineering, image processing and computer graphics.²

These descriptions, formulae and illustrations are such as to completely and clearly specify algorithms which can be implemented in a straightforward manner by programming a programmable computer imaging device such as a frame buffer.

For example, the programmable frame buffers (some with onboard special-purpose microprocessors for graphics and/or signal processing) suitable for use with personal computers, workstations or other digital computers, along with off-the-shelf assemblers, compilers, subroutine libraries, or utilities, routinely provide as standard features, capabilities which permit a user to (among other tasks): digitize a frame of a video signal in many different formats including higher-than-television resolutions, standard television resolutions, and lower-than-television resolutions, and at 8- 16- 24- and 32-bits per pixel; display a video signal in any of those same formats; change, under program control, the resolution and/or bit-depth of the digitized or displayed frame; transfer information between any of a) visible framestore memory, b) blind (non-displayed) framestore memory, and c) host computer memory, and d) mass storage (e.g., magnetic disk) memory, on a pixel-by-pixel, line-by-line, or rectangle-by-rectangle basis.³

Thus, off-the-shelf devices provide the end user with the ability to: digitize high- or low-resolution video frames; access the individual pixels of those frames; manipulate the information from those pixels under generalized host computer control and processing, to create arbitrarily processed pixels; and, display processed frames, suitable for recording, comprising those processed pixels. These off-the-shelf capabilities are sufficient to implement an image processing system embodying the information manipulation algorithms or system designs specified herein.

Similarly, higher performance and throughput (as well as higher cost and more programming effort), programmable devices, suitable for broadcast or theatrical production tasks, provide similar and much more sophisticated capabilities, including micro-coding whereby image processing algorithms can be incorporated into general purpose hardware, are available as off-the-shelf programmable systems.⁴

Additionally, specialized (graphic and image processing) programmable microprocessors are available for incorporation into digital hardware capable of providing special-purpose or general-purpose (user-programmable) image manipulation functions.⁵

Further, it is well known by those skilled in the art how to adapt processes that have been implemented as

software running on programmable hardware devices, to designs for special purpose hardware, which may then provide advantages in cost vs. performance.

In summary, the disclosure of the instant invention will focus on what is new and novel and will not repeat the details of what is known in the art.

One of the major applications intended for the instant invention is the incorporation of the algorithms disclosed herein into a film chain (a film to video transfer device). Such transfers are an important and costly part of the television motion picture industry. Much time and effort is expended in achieving desired and artistic results. And, in particular, the scene-by-scene color correction of such transfers is common practice.

Thus, in the instant disclosure, it will be suggested that practitioners make adjustments to the operational parameters of the disclosed algorithms in order to better achieve desired results. Further, it will be suggested to such practitioners that such individual adjustments may be applied to images or image portions exhibiting different characteristics.

Inventor's earlier relevant and published work, includes the following:

1. Early work in film colorization lead to the development of using shape interpolation (sometimes called image warping) and cross-dissolving, as applied to key-frame color signals, for the reduction of information storage and processing requirements.
2. Later work in film colorization and 2D to 3D conversion comprised, in part, improved methods of generating image boundary information.
3. Later work in 2D to 3D image conversion comprised, in part, the creation of 3D images by: extracting texture maps and 3D shape and motion information from motion picture sequences; and, re-applying those textures to other versions of the 3D shapes with which they were originally associated with.
4. Work in image compression and bandwidth reduction lead to the development of processes and devices for: time-varying data selection and arrangement (with improved perceptual results); off-line computation and recording for bandwidth reduction; variable pixel geometry; and, the incorporation of additional information into the blanking intervals of a frame prior to the one with which that additional information is to be associated at reception, permitting multi-frame-time and/or pipelined decoding and reintegration of that additional information.
5. A version of Inventor's paper, *STEREOSYNTHESIS: A Process for Adapting Traditional Media for Stereographic Displays and Virtual Reality Environments*, Proceedings of The Second Annual Conference on Virtual Reality, Artificial Reality, and Cyberspace, San Francisco, Meckler, 1991, provides further details on his STEREOSYNTHESIS™ 2D to 3D image conversion technology.

The following are publicly available, in the prior art, not (in and of themselves) the subject of the instant invention, and within the knowledge and familiarity of those skilled in the art.⁶

1. *Shape and Motion from Image Streams under Orthography: a Factorization Method*, Carlo Tomasi and Takeo Kanade, International Journal of Computer Vision, volume 9, number 2, pages 137-154, Kluwer Academic Publishers, The Netherlands 1992.
2. *Shape and Motion from Image Streams: a Factorization Method—Part 3: Detection and Tracking of Point Features*, Carlo Tomasi and Takeo Kanade, Carnegie Mellon University, Pittsburgh 1991.
3. *The Magic of Image Processing (Chapter 8, Morphing)*, Mike Morrison, SAMS Publishing, Indianapolis 1993.
4. Four papers from: *Computer Graphics: Proceedings of the 1992 SIGGRAPH Conference*; Volume 26, Number 2, July 1992, ACM Press, New York 1992.
 - a. *Feature Based Image Morphing*, Thaddeus Beier and Shawn Neely, at page 35.
 - b. *Scheduled Fourier Volume Morphing*, John F. Hughs, at page 43.

- c. *A Physically Based Approach to 2-D Shape Blending*, Thomas W. Sederberg and Eugene Greenwood, at page 25.
- d. *Shape Transformation for Polyhedral Objects*, James R Kent, Wayne E. Carlson and Richard E. Parent, at page 47.

5 5. *Handbook of Pattern Recognition and Image Processing (Chapter 13 A Computational Analysis of Time-Varying Images; Chapter 14 Determining Three-dimensional Motion and Structure from Two Perspective Views; and, Chapter 9 Image Segmentation)*, Ed. Tzay Y. Young, Academic Press, Inc., New York 1986.

These cites are being provided as references on: morphing; the extraction of 2D and 3D shape and motion information from motion sequences; and, the detection, creation and use of image boundaries and segments.

10 Commercial black & white and, later, color television has been available since the 1940s. American and Japanese systems offer 525 line frames, 30 times each second, while most European systems offer a higher resolution 625 line frame but run at a frame rate of 25 per second. Higher resolution military and laboratory video systems exist and, recently, a commercial high definition television standard (HDTV) has been developed to improve delivered image quality.⁷

15 In the US, motion picture film is projected at 48 frames per second (FPS) by showing each of 24 pictures twice. Recently, a system was developed by Douglas Trumbull called Showscan. It provides 60 FPS, with 60 pictures each shown only once, to improve visual quality.

When color was added to US black & white television, it was decided to adopt a "compatible" system, which enables black & white sets to receive color television signals and display them in black & white, while color
20 sets display the same signals in color. Similarly, it has been suggested that the HDTV signal be compatibly receivable by standard televisions displaying standard resolution pictures, as well as by HDTV receivers. HDTV provides both more video lines and more pixels (from PICTURE ELEMENTS: visual data points) per line. It has been suggested that the standard television channels can be used to transmit a "compatible" standard resolution signal while a second channel (not receivable by a standard television) be used to transmit the "inbetween" higher resolution information.
25 However, HDTV may also display a wider picture when compared with standard television. Inclusion of the extra "side strips" in a compatible broadcast system has been one of the main problems.

It is established practice to transmit motion picture film, which has a much higher resolution and a different frame rate, over a broadcast television channel by use of a film chain. Essentially a motion picture projector coupled to a television camera, the film chain synchronizes the two imaging systems. In newer film chain systems the video
30 camera has been replaced by a digital image sensor and digital frame store. In the US, each video frame consists of two interleaved video fields, resulting in 60 fields per second. US film runs at 24 frames per second. This results in a ratio of 2.5 video fields per film frame. Practically, this is achieved by alternating 3 repeated video fields and 2 repeated video fields for alternate film frames. The spatial resolution of the image is reduced by the characteristics of the video camera.

35 It is also established practice to generate synthetic television signals (without a camera) by using electronic devices such as character (text) generators, computer graphic systems and special effects generators.

Recent developments in home televisions and VCRs include the introduction of digital technology, such as full-frame stores and comb filters.

40 There exist many techniques for bandwidth compression of electronic signals, a number of which have been applied to television systems. These are particularly useful for transmitting images from space probes or for satellite transmission, where resources are limited.

DESCRIPTION OF INVENTION

The instant invention comprises a method, process or algorithm, and variations thereon, including motion

detection, cross-dissolving and shape interpolation; devices or systems for practicing that method; and, product (generally motion picture film, videotape or videodisc, analog or digitally stored motion sequences on magnetic or optical media, or a transmission, broadcast or other distribution of same) produced by the method and/or system.

The purpose to which the invention is applied is to process (generally by digital computer image processing) a motion picture sequence in order to produce a processed motion picture sequence which exhibits: an increase in the perceived quality of that sequence when viewed; and/or a decrease of the requirements for information storage or transmission resources without significantly effecting image quality (i.e., data compression or bandwidth reduction).

In order to understand the invention more fully, it is helpful to examine certain aspects of film and video display systems, their shortcomings, and the functioning of the human visual system. The reader is directed to consult the parent application, of which the instant application is a continuation-in-part, for further details.

SPATIAL/TEMPORAL CHARACTERISTICS OF FILM AND VIDEO SYSTEMS:

Film and video display systems each have their own characteristic "signature" scheme for presenting visual information to the viewer over time and space. Each spatial/temporal signature (STS) is recognizable, even if subliminally, to the viewer and contributes to the identifiable look and "feel" of each medium.

Theatrical film presentations consist of 24 different pictures each second. Each picture is shown twice to increase the "flicker rate" above the threshold of major annoyance. However, when objects move quickly, or contrast greatly, a phenomenon known as strobing happens. The viewer is able to perceive that the motion sequence is actually made up of individual pictures and motion appears jerky. This happens because the STS of cinema cameras and projectors is to capture or display an entire picture in an instant, and to miss all the information that happens between these instants.

In cinematography, the proportion of time the shutter is open during each 1/24th second can be adjusted. Keeping the shutter open for a relatively long time will cause moving objects to blur. In "stop motion" model photography it is now common practice to leave the shutter open while the model is moved for each exposure, rather than to take a series of static images (the technique, first popularized at Industrial Light and Magic, is referred to as "go motion" photography). In both cases, each motion picture frame is taken over a "long" instant, while objects move. This does cause motion blurring, but does also lessen the perception of strobing; the "stuttering" nature of the film STS has been lessened by temporal smearing.

A phenomenon related to strobing, which also is more noticeable for contrasty or fast moving situations, is called doubling. As noted, each motion picture frame is shown twice to increase the flicker rate. Thus, an object shown at position A in projected frame 1, would again be shown at position A in projected frame 2, and would finally move to position B in projected frame 3. The human eye/brain system (sometimes called the Retinex, for RETinal-cerebral complex) expects the object to be at an intermediate position, between A and B, for the intermediate frame 2. Since the object is still at position A at frame 2, it is perceived as a second object or ghost lagging behind the first; hence, doubling. Again, this is a consequence of the STS of film projection. The overall result is a perceived jitteriness and muddiness to motion picture film presentations, even if each individual picture is crisp and sharp.

Video, on the other hand, works quite differently. An electron beam travels across the camera or picture tube, tracing out a raster pattern of lines, left-to-right, top-to-bottom, 60 times each second. The beam is turned off, or blanked, after each line, and after each picture, to allow it to be repositioned without being seen.

Except for the relatively short blanking intervals, television systems gather and display information continuously, although, at any given time, information is being displayed for only one "point" on the screen. This STS is in marked contrast to that of film. Some defects of such a system are that the individual lines (or even dots)

of the raster pattern may be seen because there is only a limited number of individual dots or lines — i.e., resolution — that can be captured & displayed within the time or bandwidth allotted to one picture.

In US commercial television systems, each 1/30 second video frame is broken into two 1/60 second video fields. All the even lines of a picture are sent in the first field, all the odd lines in the second. This is similar to showing each film frame twice to avoid flickering but here it is used to prevent the perception of each video picture being wiped on from top to bottom. However, since each video field (in fact each line or even each dot) is scanned at a different time, there is no sense of doubling.

The muddiness or opacity of film presentations, when compared to video, is related to the repeated presentation of identical information to the human visual system. This can be demonstrated by watching material transferred from film to video using newer equipment. As explained above, each film frame is repeated for either 3 or 2 video fields during transfer. Newer film chains can pan, pull or tilt across the visual field during transfer. In doing so, each video field contains unique information. Even if the same film frame is scanned, it is scanned from a different position or orientation. During those brief sequences when a camera move is added by the film chain equipment, there is a perceivable increased clarity to the scene.

In summary, film systems deal with information everywhere at once, but for only small slices of time. Television systems deal with information (almost) all the time, but for only small slices of space. Each STS approach leads to characteristic perceivable anomalies or artifacts; primarily, temporal muddiness for film, low geometric resolution for video.

The instant invention can employ motion detection and/or interpolative techniques to create an STS scheme which will reduce both types of perceivable anomalies and which can be used to reduce the bandwidth required to transmit image motion sequence signals.

THE INVENTION IN BRIEF:

The basis of the instant invention is that the human visual system responds better to information display systems that present unique information at each frame. Standard theatrical motion picture films provide only 24 unique images of 48 presented each second. On the other hand, standard broadcast television (not originated on film) provides 60 unique field images each second, but at lower resolution; and, Showscan provides both high temporal and high geometric resolution.

The instant invention will employ high-level algorithms and system designs to process motion picture sequences (originating in film, video or otherwise) to produce film, video or digital presentations that meet the uniqueness requirement. This will be done by synthesizing information frames for times intermediate to those available. The lower-level algorithms involved include motion detection and specification, image segmentation, shape interpolation and cross-dissolving. The last two, in combination, are sometimes referred to as "transition image morphing".⁸

In many embodiments, this processing will be applied to a source image stream to create a processed image stream by the application of much computation and, optionally, some human intervention and assistance. The results can be recorded (perhaps, in an off-line manner) and then distributed via any standard information delivery method, or as any standard information product. In particular, the processing of images derived from standard theatrical motion picture film at 24 FPS to produce video (or film) at 60 FPS is envisioned as an improved film chain device.

In addition, since a higher-frame rate image stream can be created, from a lower-frame rate image stream, some embodiments will permit a reduced-frame rate image stream to be transmitted (or stored), generally with additional motion specification information, and a higher-frame rate image stream constructed at the reception (or access) and display site. Thus, a data compression or bandwidth reduction will result with this embodiment which may be used to reduce storage or transmission requirements, or can be used to make way for information additional

to the image stream which can comprise: additional resolution or definition; additional image area (e.g., wide-screen side-strips); 3D information in the form of a second image, or from which two images can be created by combination with the first; interactive r game data; hyper- or multimedia data; image segmentation data showing areas of motion or where different algorithms are to be applied; or, the interleaving f several program channels.

In particular, it is noted that, in addition to standard television broadcasting, such compression is very desirable for a number of other applications. Specifically: so-called "500 channel" cable (or via satellite broadcast, fiber or phone line) television; digital image streams to be displayed from computer disk or CD-ROM; image streams via communication lines for on-line multimedia or video conferencing; storage of video signals on analog or digital tape (or other magnetic or optical media); the transmission of HDTV, stereographic television, or new "digital" television signals.

DETAILED DESCRIPTION WITH DRAWINGS

What follows is a detailed description with drawings that will illustrate several preferred embodiments of the instant invention.

Referring, first, to Table I, below, note that: film frame 0 exactly corresponds in time with an even video field 0; film frame 1 falls between even video field 2 and odd video field 3; film frame 2 exactly corresponds in time with an odd video field 5; film frame 3 falls between odd video field 7 and even video field 8; and, film frame 4 exactly corresponds in time with an even video field 10, starting the repeat of the 1/6th second temporal cycle.

* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	1/120 SECOND CLICKS
e	o	e	o	e	o	e	o	e	o	e	o	VIDEO FIELD TYPE
0	1	2	3	4	5	6	7	8	9	10		VIDEO FIELD COUNT
0		1		2		3		4				FILM FRAME COUNT

TABLE I: Temporal Alignment of Film Frames and Video Fields

In the tersest terms, the basic embodiment of the invention will be to use shape interpolation and cross-dissolving (i.e., a process akin to image morphing) to derive, from pairs of film images, intermediate images, for the purpose of presenting unique and temporally appropriate images at each video field.

Table II shows the setting of the morph parameter (0% to 100%) and which film images are used to create each video field. Note that a morph parameter of 100% corresponds to using the first of the two film frames alone and unprocessed. Similarly a morph parameter of 0% would correspond (if used) to using the second of the two film frames alone and unprocessed. The number in parenthesis is the complementary percentage from the perspective of the second frame.

<u>First Film Frame</u>	<u>Second Film Frame</u>	<u>"Morph" Parameter</u>	<u>Video Field</u>
0	1	100% (0%)	0 e
0	1	60% (40%)	1 o
0	1	20% (80%)	2 e
1	2	80% (20%)	3 o
1	2	40% (60%)	4 e
2	3	100% (0%)	5 o
2	3	60% (40%)	6 e

2	3	20% (80%)	7 o
3	4	80% (20%)	8 e
3	4	40% (60%)	9 o
4	5	100% (0%)	10 e temporal
			repeat
			here

TABLE II: Morphing Parameter for Film Frames and Video Fields

As shown, the image data is derived from the film frames. For interpolation, shape data is also required. This may be provided by a computer/human collaborative system such as that disclosed by Inventor for film colorization or 2D to 3D image conversion (or as used by PDI). Please refer to Figures from Inventor's earlier patents and applications for system diagrams; only the particular software and algorithms being run will change. As subsequently disclosed by Inventor, such systems can also be made to work in a more or less automatic fashion by the incorporation into the system of additional software capabilities to extract image boundary (segmentation) information and/or motion data. Similarly, those capabilities may be applied here to generate boundary information that may be used to implement the morphing functions.

Such automatic operation was considered less than optimal for Inventor's earlier systems because it was necessary to identify and separate actual objects from within the frame. At least for some morphing algorithms, it is only necessary to identify the areas of the image that move (irrespective of whether those areas correspond to real-world coherent objects) or which need be associated from key frame to key frame. Further, the difference between one film frame and the next (within a scene) are generally quite small. In contrast, Inventor's film colorization system employed key frames many film frames apart. Therefore, the use of automatic boundary extraction (particularly based on motion) and motion analysis algorithms will provide change information appropriate to the close in time "micro-morphing" task at hand.

In particular, a technique that extracts "optical flow" will be used as follows. There, rather than boundary information, what is extracted is a field showing how the various areas (e.g., individual pixels) of the image are moving (both magnitude and direction) from frame to frame. This information may include translation, sizing, skewing or rotation changes. See Figure 1. Additionally, pixels may "appear" or "disappear" as object rotates and new areas come from behind or old areas go out of view. Similarly, as objects mutually intersect, portions may become newly visible or obscured. See Figure 2.

This optical flow data can be used in lieu of the interpolated boundaries to provide the warping aspect of a morphing like function, with an interpolated field function applied to the pixels of the entire frame, pixel-by-pixel. In particular, optical flow or other motion data may be provided over the entire image or only at selected points (e.g. on a regular grid). See Figure 3. The data can then be interpolated between those points given, to arrive at appropriate values for each pixel in the image. For embodiments where this data will have to be transmitted (see below) data may be sent only for certain of the points in each frame. Those points with the most significant data may be sent, or a more regular parsing may be employed. For a simple example, if one considers a checkerboard overlaid on such a grid, the "black points" may be alternated with the "white points". At each frame the data of the more current set will be given heavy weight; however, the points sent for prior or subsequent frames may also be consulted (perhaps averaged over time) but, perhaps with less weight. Alternately, a more complex "variable STS" type of pattern may be employed to select which points to transmit (or the position of those points sent) with each frame.

Whichever technique is employed for image warping, the percentages of Table II are applied to that process, as well as the cross-dissolving function, and unique frames are created for each video field (or for additional film

frames).

The above will be accomplished either automatically or with human operator participation; but, in many embodiments (particularly where optical flow computations are being used to compute motion for image warping) the process will be accomplished in an off-line manner. That is, the image analysis and processing computations will be done on a frame-by-frame basis (although, particularly for the analysis, several frames will be "considered" simultaneously) and these frames will be created, collected and committed to film or videotape on a slower than real-time basis.

For other embodiments, the motion/change/shape data calculations will be performed but, rather than producing the new frames, the old frames and the motion data will be recorded or transmitted. Upon access or reception, the low-frame rate image data and motion data will be combined, in real time, to create a full-frame rate image stream. The advent of very-high-performance consumer electronics (e.g., interactive game settop boxes and the like) will provide a hardware environment within which such computations may be carried out. See Figure 4. Pipelined architecture and variable geometry frame stores (as disclosed in Inventor's other applications) will be useful to implement such devices. Further, for such real-time applications, computationally simpler embodiments will be preferred.

Eventually, settop boxes and the like, may become available which can, in real time, perform the entire process (motion analysis and morphing). Until that time, both image and motion data will have to be delivered and utilized. Several embodiments of how to accomplish this follow.

In a straightforward embodiment, image data frames may be alternated with shape or motion data. And that shape or motion data may be associated with the previous image data, the later image data or "inbetween the two". See Figure 5.

If shape data are used, the shapes are interpolated between shape data frames.

If motion data are used, the motion offsets may be applied in several ways. If a motion offset data frame is supplied, it can represent a 1/120th second change. Thus, for a video field at or after the time of the film image: for a 100% morph parameter the offset is not applied since the image is used unchanged; for an 80% parameter it is applied once; for a 60% parameter it is applied twice (in succession or twice as strongly); for a 40% parameter it is applied three times; for a 20% parameter it is applied four times; for a 0% parameter it is not applied since the image is not used.

Similarly, for a video field at or before the time of the film image: for a (100%) morph parameter the offset is not applied since the image is used unchanged; for a (80%) parameter it is applied once but with a reversed sign; for a (60%) parameter it is applied twice (in succession or twice as strongly) but with a reversed sign; for a (40%) parameter it is applied three times but with a reversed sign; for a (20%) parameter it is applied four times but with a reversed sign; for a (0%) parameter it is not applied since the image is not used.

Alternately, the shape or motion frame may be considered to be "between" the image frames. Then the same shape/motion data frame will be applied to the image frames on either side, but in opposite directions. If an image frame is the first of the pair the shape/motion frame to the right is applied with positive sign; if an image frame is the second of a pair, the shape/motion frame to the left is applied with negative sign. See Figure 6.

With either shape interpolation or motion offset application, if only two shape or motion data frames are applied a linear interpolation between the two is possible.

However, for more sophistication, the values from one or more frames before and or after the frame (or frame pair) in question can be consulted. Thus, curve fitting algorithms (e.g., splines) can be applied to all data dimensions (translations in X and Y, rotations, skews, size changes, sources or sinks; or more with 3D shape/motion data). In this way, more natural and sophisticated changes, that progress non-linearly, from frame to frame, can be computed. See Figure 7 for examples shown for a single parameter.

By the method described above, film may be stored as, or sent via, video with some additional information space left available. For example with five video fields used to hold two film images, two fields may be applied to each film frame, with the two shape/motion data frames contained in the fifth field. However, the shape/motion data can, instead, be put in the blanking intervals of those frames (or, as disclosed for side strip information in Inventor's co-pending application, in a previous frame) leaving one field free. Further, by applying line doubling interpolation (this can be tolerated since full-frame video provides much better response vertically than horizontally) only one frame each of the two frames need be sent, and then three of five video fields can be made available. See Figure 8. These additional fields (comprising as much as 60% of the image stream) may be used for: additional resolution or definition (in both directions or in bit-depth); additional image area (e.g., HDTV, wide-screen or "letterbox" side-strips); 3D information in the form of a second image, or from which two images can be created by combination with the first; interactive or game data; hyper- or multimedia data; image segmentation data showing areas of motion or where different algorithms are to be applied; or, the interleaving of several program channels. The specifics of these uses will not be disclosed here, some have already been disclosed by Inventor in other applications or patents. The details of such use, in general, are not in and of themselves considered the substance of the present invention (except where specific novel details are provided); however, the application of the "morphing" frame creation process, and the ensuing "freeing up" of video bandwidth, resulting in these possible uses, is the substance of the present invention.

As explained, above, system diagrams for the instant invention are virtually identical to those provided by Inventor in earlier applications, for either computer assisted or automatic systems. However, an information or software flow diagram is provided as Figure 9.

Next, a more sophisticated embodiment is described, which will be particularly useful where pixel sinks and sources occur, and which was also described in Inventor's earlier applications and publications in order to create "Virtual Reality" presentations based on films.

In this case:

1. Image analysis algorithms are first applied to the image sequence to extract 3D shape and motion data.
2. The bitmaps representing the surfaces of these objects are extracted from the image and the inverse of the projection transform is used to "unwrap" the surface images from the 3D shapes derived in step 1 to create texture maps for each 3D object. These may be pieced together from several images either up- or down-stream of the frame in question.
3. Based on the 3D motion data extracted in step 1, intermediate 3D frame scenes are created repositioning or reshaping each 3D object.
4. For each object, texture maps from source images, on either side of the intermediate frame to be created, are cross-dissolved (or the closest texture map may be used).
5. The texture maps are then reapplied to the distorted and/or repositioned 3D objects and 2D projections (or stereoscopic pairs of 2D projections) are created as intermediate frames.

See Figure 10.

The above may be used as an alternative to the 2D embodiments which came before, or aspects of each embodiment may be combined. It is less likely that this 3D embodiment will be usable in a completely automatic fashion, and less likely still that it may be used for a real-time system (at least with current commercial level technology). Nevertheless, for processing 24 FPS theatrical motion picture film for 60 FPS projection f video transfer, these techniques may be useful to process problematic scenes not adequately handled by other methods.

These techniques may be combined with other data reduction techniques to advantage. For example, using image segmentation data described elsewhere, data may be sent/stored, in addition to image frames and shape/motion frames, so that various areas of frames in a sequence may be assembled from several methods.

For example (see Figure 11):

1. Some areas may be retained from one frame to the next. Particularly since analysis of motion data will be an important aspect of the basic instant invention areas that lack motion or change will be detected. Thus, part of the data sent can include (or be deduced from the motion data sent) a map of area that move so little that then need not be updated for at least the current frame.
2. Some areas may change so drastically that the present invention will not prove adequate and, for those areas (also indicated by some {presumably highly compressed} area map) replacement data would be sent which may be compressed by any compatible data compression technique now extant or later developed.
3. Those areas remaining may be interpolated by the techniques disclosed herein.

The flows depicted in the software flow diagrams herein are exemplary, some items may be ordered differently, combined in a single step, skipped entirely, or accomplished in a different manner. However, the depicted flows will work. In particular, some of these functions may be carried out by hardware components, or by software routines residing on, or supplied with, such a component.

Similarly the systems depicted in the system diagrams herein are exemplary, some items may be organized differently, combined in a single element, split into multiple elements, omitted entirely, or organized in a different manner. However, the depicted systems will work. In particular, some of these functions may be carried out by hardware components, or by software routines residing on, or supplied with, such a component.

It will thus be seen that the objects set forth above, among those made apparent from the preceding description, are efficiently attained and certain changes may be made in carrying out the above method and in the construction set forth. Accordingly, it is intended that all matter contained in the above description or shown in the accompanying figures shall be interpreted as illustrative and not in a limiting sense.

While there has been shown and described what are considered to be preferred embodiments of the invention, it will, of course, be understood that various modifications and changes in form or detail could readily be made without departing from the spirit of the invention. It is, therefore, intended that the invention be not limited to the exact form and detail herein shown and described, nor to anything less than the whole of the invention herein disclosed as hereinafter claimed.

I claim:

NOTES

1. Typical examples include:

Digital Video: Selections from the SMPTE Journal and Other Publications, Society of Motion Picture and Television Engineers, Inc. (SMPTE), 1977.

Digital Video Volume 2, SMPTE 1979.

Digital Video Volume 3, SMPTE 1980.

Graphics Engines, Margery Conner, Electronic Design News (EDN), Cahners Publishing Company, Newton, MA, Volume 32, Number 5, March 4, 1987, pages 112-122.

Algorithms for Graphics and Image Processing, Theo Pavlidis, Computer Science Press 1982.

Computer Vision, Ballard and Brown, Prentice-Hall, Englewood Cliffs 1982.

Industrial Applications of Machine Vision, IEEE Computer Society, Los Angeles 1982.

Structured Computer Vision, Ed. Tanimoto and Klinger, Academic Press, New York 1980.

Computer Architecture for Pattern Analysis and Image Database Management, IEEE Computer Society Press, Hot Springs 1981.

Image Processing System Architectures, Kittler & Duff, John Wiley & Sons, Inc., New York 1985.

Multiresolution Image Processing and Analysis, Ed. A. Rosenfeld, Springer-Verlag, New York 1984.

Image Reconstruction from Projections, Gabor T. Herman, Academic Press 1980.

Basic Methods of Tomography and Inverse Problems, Langenberg and Sabatier, Adam Hilger, Philadelphia 1987.

US Patent Number 2,940,005 issued June 7, 1960, Inventor: P. M. G. Toulon.

Principles of Interactive Computer Graphics, Second Ed., Newman & Sproull, McGraw-Hill Book Company, New York 1979.

Advances in Image Processing and Pattern Recognition, Elsevier Science Publishers B.V., Amsterdam, 1986.

Image Recovery Theory and Application, Henry Stark, Academic Press, Inc., New York 1987.

Handbook of Pattern Recognition and Image Processing, Ed. Tzay Y. Young, Academic Press, Inc., New York 1986.

- Fundamentals of Interactive Computer Graphics*, Foley and Van Dam, Addison-Wesley, New York, 1982.
- Real Linear Algebra*, Anatal E. Fekete, Marcel Dekker, Inc., New York 1985.
- Finite Dimensional Multilinear Algebra, Parts I & II*, Marvin Marcus, Marcel Dekker, Inc., New York 1973.
- Sparse Matrix Computations*, Ed. Bunch & Rose, Academic Press, Inc., New York 1976.
- Matrix Computations and Mathematical Software*, John R. Rice, McGraw-Hill Book Company, New York 1981.
- The Architecture of Pipelined Computers*, Peter M. Kogge, McGraw-Hill Book Company, New York 1981.
- Digital System Design and Microprocessors*, John P. Hayes, McGraw-Hill Book Company, New York 1984.
- Digital Filters and the Fast Fourier Transform*, Ed. Bede Liu, Dowden, Hutchinson and Ross, Inc., Stroudsburg 1975.
- Hardware and Software Concepts in VLSI*, Ed. Guy Rabbat, Van Nostrand Reinhold Company, Inc., New York 1983.
- Digital Signal Processing*, Oppenheim and Schaffer, Prentice Hall, Inc., Englewood Cliffs 1975.
- Movements of the Eyes*, R. H. S. Carpenter, Pion, Limited, London 1977.
- Service Manual: DCX-3000 3-Chip CCD Video Camera*, SONY Corporation.
- Color Television: Principles and Servicing* 1973.
- Multi-Dimensional Sub-Band Coding: Some Theory and Algorithms*, Martin Vetterli, Signal Processing 6 (1984) 97-112, Elvsevier Science Publishers B.V. North-Holland, p. 97-112.
- The Laplacian Pyramid as a Compact Image Code*, Burt and Adelson, IEEE Transactions on Communications, Vol. Com-31, No. April 1983, p. 532-540.
- Exact Reconstruction Techniques for Tree-Structured Subband Coders*, Smith & Barnwell, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-34, No. 3 June 1986, p. 434-441.
- Theory and Design of M-Channel Maximally Decimated Quadrature Mirror Filters with Arbitrary M, Having the Perfect Reconstruction Property*, P.P. Vaidyanathan, IEEE Transactions on Acoustics, Speech and Signal Processing, V 1. ASSP-35, No. 4, April 1987, p. 476-492.
- Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes*, Esteban & Galand,

IBM Laboratory, 06610, La Gaudé, France.

Extended Definition Television with High Picture Quality, Broder Wendland, SMPTE Journal, October 1983, p. 1028-1035.

2. See, for example:

Digital Video: Selections from the SMPTE Journal and Other Publications, Society of Motion Picture and Television Engineers, Inc. (SMPTE), 1977.

Digital Video Volume 2, SMPTE 1979.

Digital Video Volume 3, SMPTE 1980.

Extended Definition Television with High Picture Quality, Broder Wendland, SMPTE Journal, October 1983, p. 1028-1035.

Computer Graphics: Proceedings of the 1992 SIGGRAPH Conference; Volume 26, Number 2, July 1992, ACM Press, New York 1992.

3. For example:

PIP-512, PIP-1024 and PIP-EZ (software); PG-640 & PG-1280; MVP-AT & Imager-AT (software), all for the IBM-PC/AT, from Matrox Electronic Systems, Ltd. Que., Canada.

The Clipper Graphics Series (hardware and software), for the IBM-PC/AT, from Pixelworks, New Hampshire.

TARGA (several models with software utilities) and AT-VISTA (with software available from the manufacturer and Texas Instruments, manufacturer of the TMS34010 onboard Graphics System Processor chip), for the IBM-PC/AT, from AT&T EPICenter/Truevision, Inc., Indiana.

The low-end Pepper Series and high-end Pepper Pro Series of boards (with NNIOS software, and including the Texas Instruments TMS34010 onboard Graphics System Processor chip) from Number Nine Computer Corporation, Massachusetts.

4. For example:

FGS-4000 and FGS-4500 high-resolution imaging systems from Broadcast Television Systems, Utah.

911 Graphics Engine and 911 Software Library (that runs on an IBM-PC/AT connected by an interface cord) from Megatek, Corporation, California.

One/80 and One/380 frame buffers (with software from manufacturer and third parties) from Raster Technologies, Inc., Massachusetts.

Image processing systems manufactured by Pixar, Inc., California.

And many different models of graphic-capable workstations from companies such as SUN and Silicon Graphics, Inc., including the Indy, Indigo and ONYX series.

5. For Example:

GMP VLSI Graphics Microprocessor from Xtar Electronics, Inc., Illinois.

Advanced Graphics Chip Set (including the RGB, BPU, VCG and VSR) from National Semiconductor Corporation, California.

TMS34010 Graphics System Processor (with available Software Development Board, Assembly Language Tools, "C" Cross-Compiler and other software) from Texas Instruments, Texas.

6. Other useful references include, for example:

The Interpretation of Visual Motion, Ullman, MIT Press, Cambridge 1992.

Processing Differential Image Motion, Rieger and Lawton, Journal Optical Society of America, Vol 2, No. 2, February 1985.

On the Sufficiency of the Velocity Field for Perception of Heading, Warren, Blackwell, Kurtz, Hatsopoulos and Kalish, from Biological Cybernetics, Springer-Verlag 1991.

Numerical Shape from Shading and Occluding Boundaries, Ikeuchi and Horn, Artificial Intelligence 17, North-Holland Publishing Company 1981.

Processing Translational Motion Sequences, Lawton, Computer Vision Graphics and Image Processing 22, Academic Press, Inc. 1981.

The Interpretation of a Moving Retinal Image, Longuet-Higgins and Prazdny, Proceedings of the Royal Society of London 1980.

Object Recognition by Affine Invariant Matching, Lamdan, Schwartz and Wolfson, IEEE 1982.

Sight and Mind, Kaufman, Oxford Press, New York 1974.

Perception: An Applied Approach, Schiff, Copley Publishing Group, Acton 1990.

CLAIMS

- 5 1. A method for achieving increased visual quality by converting a first image sequence of a first frame rate to a second image sequence of a second, higher, frame rate by applying shape interpolation and cross-dissolving to pairs of source images to create intermediate images and combining at least some of said first image sequence with said intermediate images.
- 10 2. A method as in claim 1, wherein said first image sequence is at 24 frames per second and said second image sequence is at 60 frames (fields) per second.
- 15 3. A product created by the process of claim 2 and recorded on film.
4. A product created by the process of claim 2 and recorded on videotape.
- 20 5. A product created by the process of claim 2 and distributed via an information bearing medium.
- 25 6. A process for data reduction whereby an image sequence to be displayed at a second, higher, frame rate is stored or transmitted as an image sequence of a first, lower, frame rate interspersed with shape/motion data frames.
- 30 7. A process for image display of the data reduced image sequence of claim 6 whereby said image sequence of a first, lower, frame rate, is processed, along with said interspersed shape/motion data to produce intermediate frames which are displayed in combination with said first image sequence.
- 35 8. An improved process for image compression whereby, for each frame in an image display sequence, the next image in said image display sequence is constructed by some combination of: retaining some portion of the previous frame in said image sequence; replacing some portion of the previous frame with new image data for said next frame; and, creating, by shape interpolation between said previous frame and a subsequent frame, some portion of said next frame.
- 40 9. A product created by conveying on an information bearing medium the information created by the process of claim 6.

OPTICAL FLOW EXAMPLES



TRANSLATION



SIZING



SKEWING



ROTATION

FIGURE 1

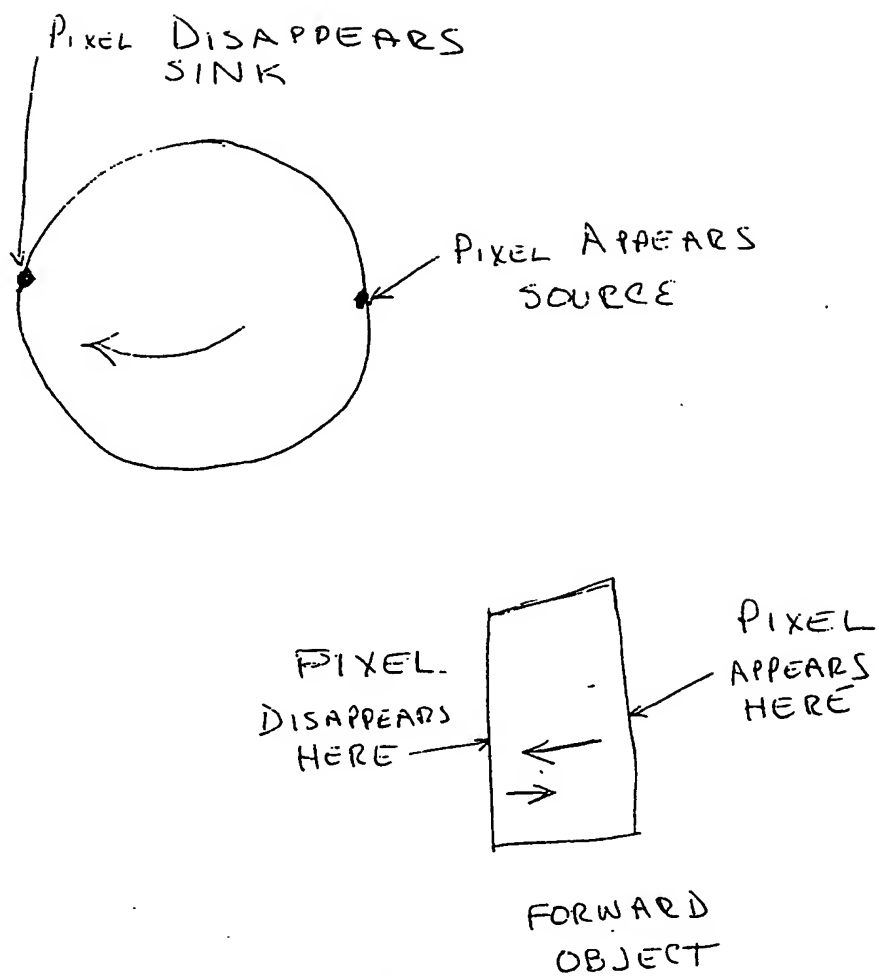


FIGURE 2

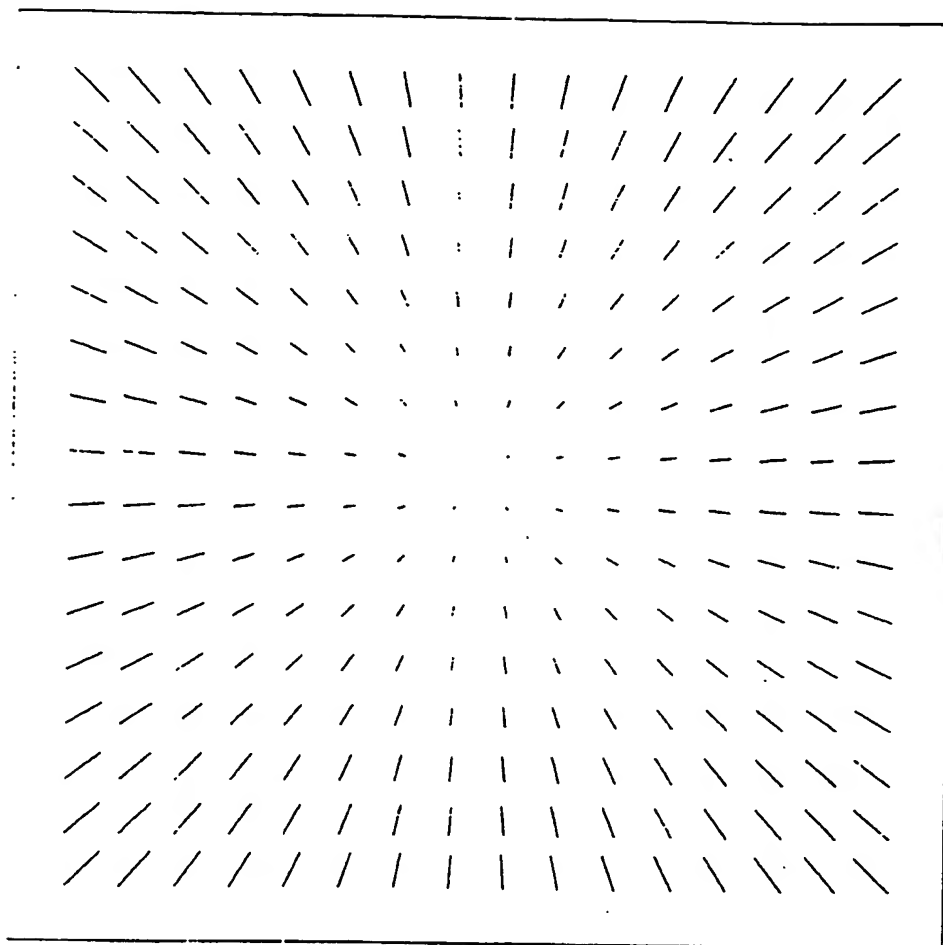
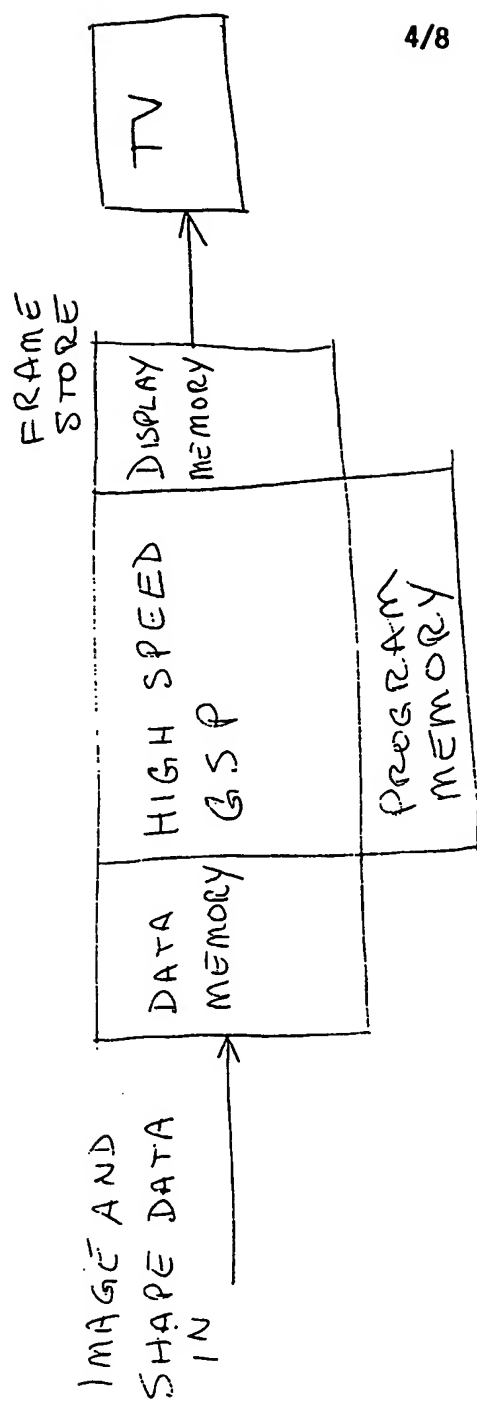


FIGURE 3

FIGURE 4

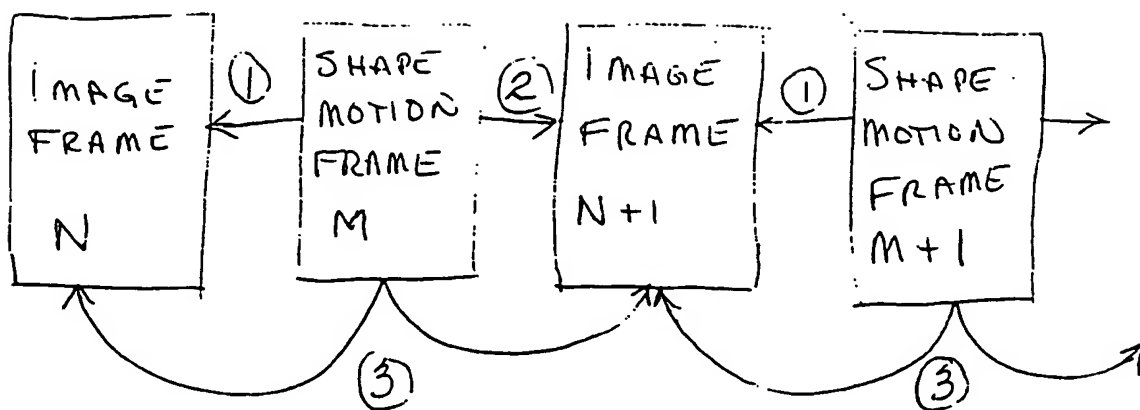


FIGURE 5

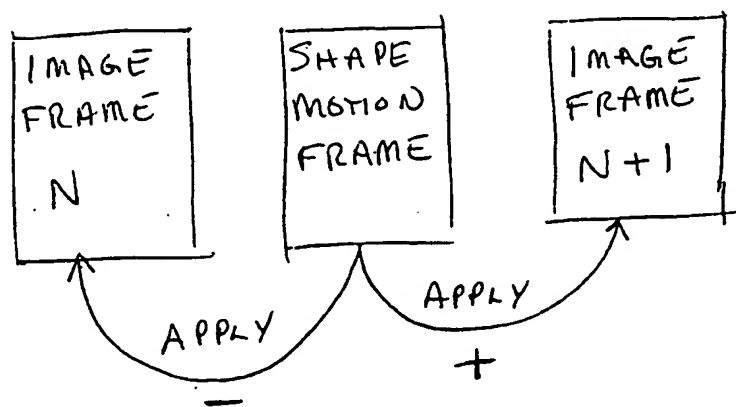


FIGURE 6

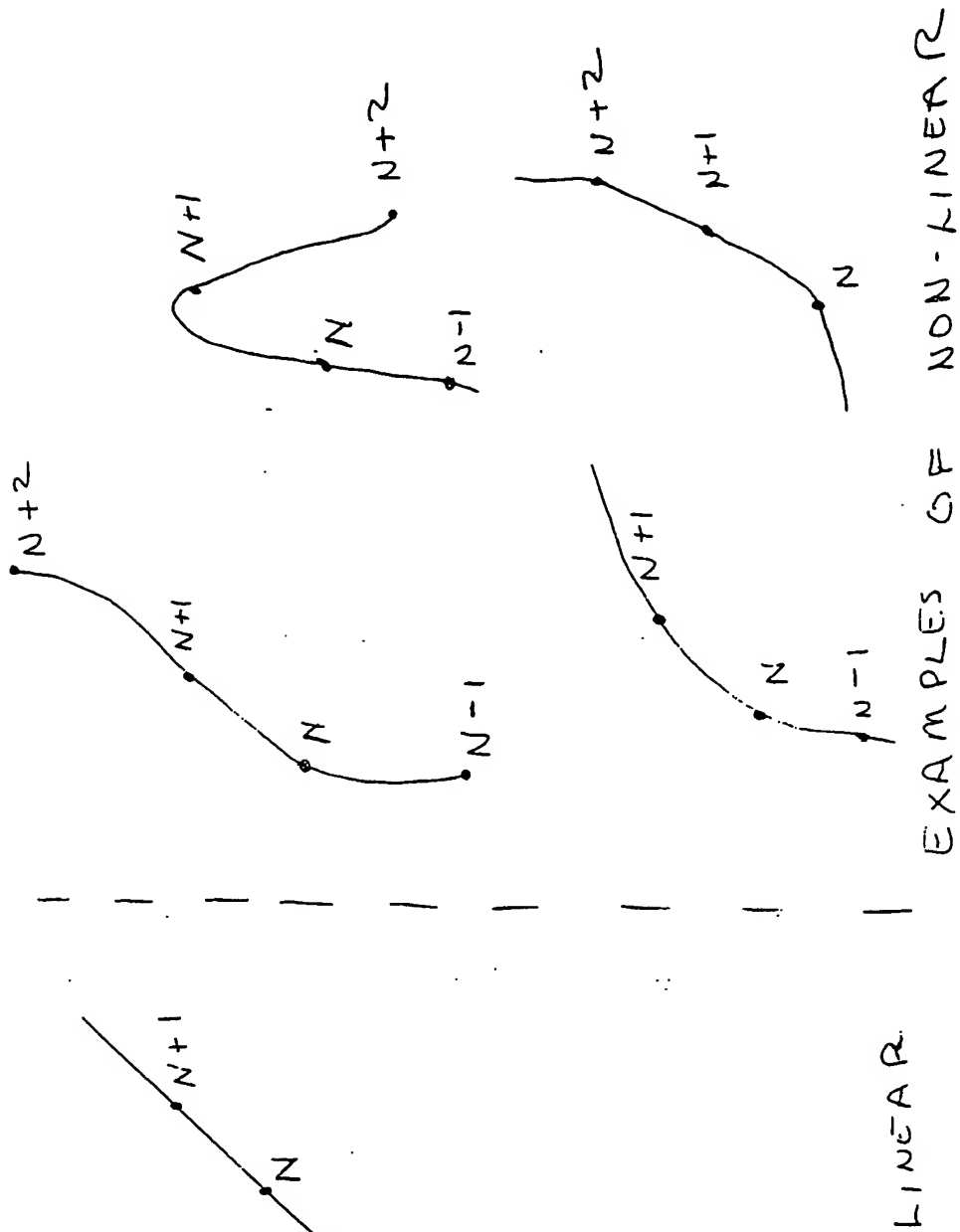


FIGURE 7

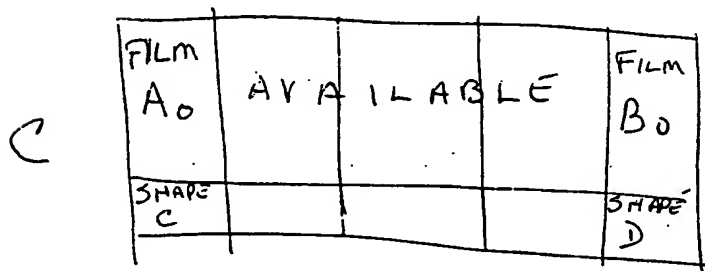
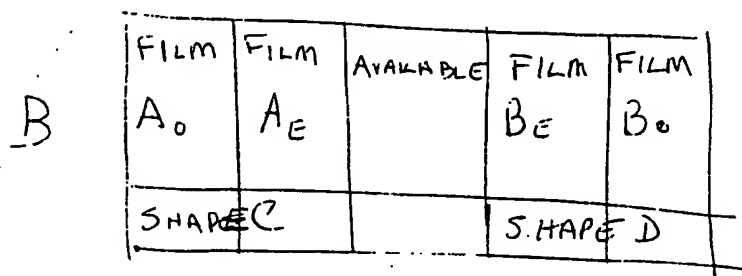
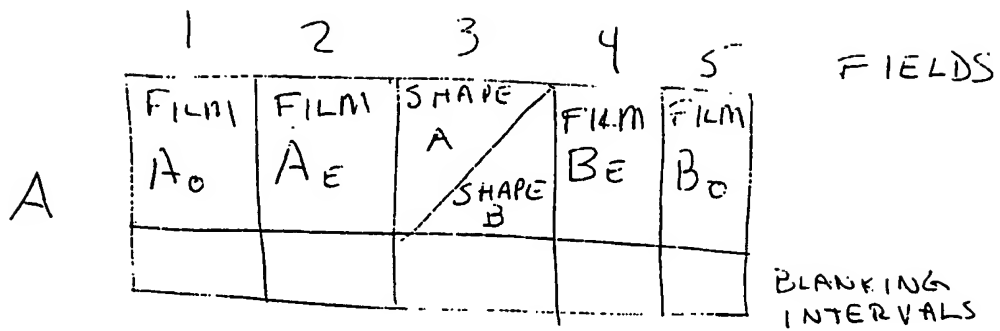


FIGURE 8

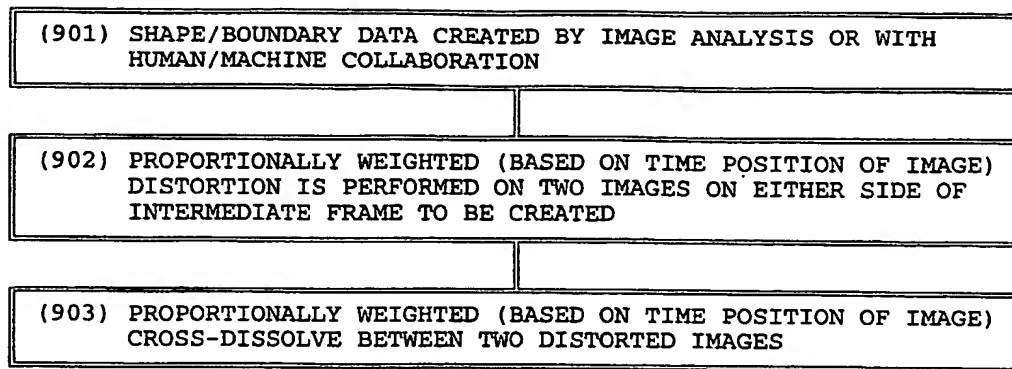


FIGURE 9: Software Flow Diagram of 2D Interpolation Embodiment

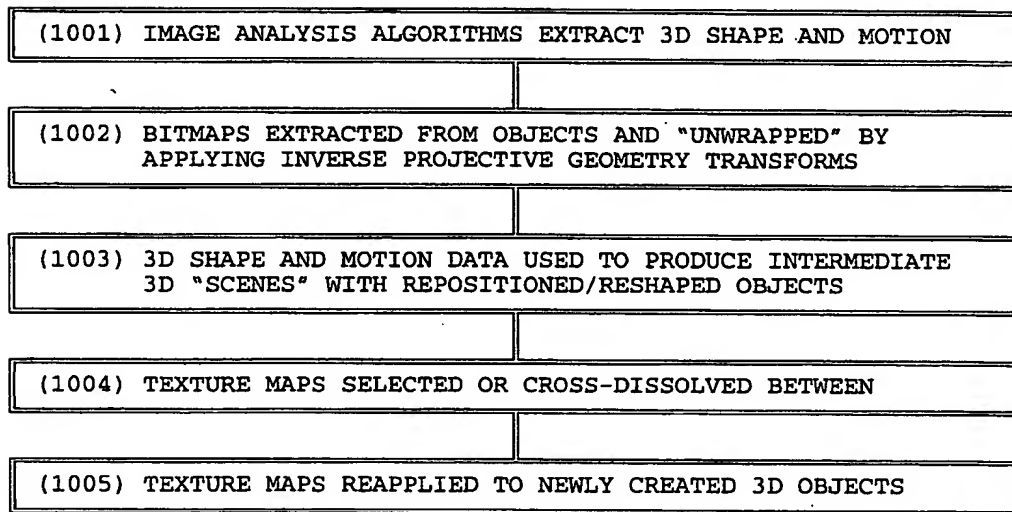


FIGURE 10: Software Flow Diagram of 3D Interpolation Embodiment

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/09813**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) :H04N 7/01

US CL :348/441, 459

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 348/441, 459, 97, 448, 453, 578; 382/308, 268, 154

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US, A, 5,410,358 (SHACKLETON ET AL) 25 April 1995, col. 3, line 5 - col. 4, line 39.	8 and 9

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	* T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
* A document defining the general state of the art which is not considered to be part of particular relevance	* X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
* E earlier document published on or after the international filing date	* Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
* L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	* & document member of the same patent family
* O document referring to an oral disclosure, use, exhibition or other means	
* P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

18 SEPTEMBER 1996

Date of mailing of the international search report

11 OCT 1996

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

VICTOR R. KOSTAK

Telephone No. (703) 305-4374